

基于人工智能的主观题自动评分算法实现

李学俊

(西南科技大学 计算机科学与技术学院, 四川 绵阳 621010)

摘 要:针对网络考试系统中主观题自动评分面临的困难和问题,将人工智能中中文自然语言理解的研究成果(空间向量模型、分词算法等)应用于网络考试中对主观题答案(包括标准答案和考生答案)的“理解”过程中,并在此基础上提出了将基于矢量空间模型的文本矢量特征匹配算法应用于计算机对考生答卷进行自动评判的过程中。实验测试数据表明,所提出的算法具有一定应用前景。

关键词: 自动评分;中文分词算法;矢量空间模型;文本匹配

中图分类号: TP 391

文献标识码: A

文章编号: 1671 - 7147(2009)03 - 0292 - 04

Algorithm Realization of the Subjective Question's Auto-Marking Based on the Artificial Intelligence

LIXue-jun

(School of Computer Science and Technolog, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract:Owing to the difficulties and problems faced by the subjective question's auto marking system in the online examination system, the paper makes use of the studied results(such as Vector Space Model (VSM), Word Segmentation algorithm and so on) of the native language understanding in the Artificial Intelligence applied in processing the subjective question's answer (including the standard answer and the student's answer), and uses the text_charactered vectormatching algorithm to auto mark those student's examination paper by the computer system. Based on the data of testing experiment, it can be seen that the algorithm has some valuable domains of application

Key words: auto-marking, word segmentation algorithm, vector space model, text matched

计算机技术和互联网技术的迅猛发展导致了传统教育模式的改变,网络教学已经成为全球性的大趋势。网络考试系统是网络教学不可或缺的重要部分,研究网络考试系统有利于实现考试的规范化、科学化、系统化。计算机可以很好地完成客观题(如选择题、判断题)的判分工作,其判分策略、关键技术及其应用实例详见文献[1]~文献[5],亦即

把考生答题结果和题目标准答案进行精确匹配,从而得出考生的得分。文献[6]提出了一种近似串匹配算法来对文本录入题进行自动评分,其本质还是进行文本的比较,与客观题的判分原理基本相同。

计算机自动评分是指利用计算机程序模拟人工评分的标准和内部过程,对开放式题目和表现性活动进行评分及反馈。对客观题的判分是通过把试

收稿日期:2009 - 03 - 01; 修订日期:2009 - 03 - 26。

基金项目:西南科技大学继续教育研究与发展基金项目(09ZW1002)。

作者简介:李学俊(1975 -),男,四川资阳人,讲师,工学博士。主要从事信息系统及信息技术、数据库及数据挖掘、多源图像融合等研究。Email:lixuejunmail@163.com

题的标准答案与考生的答案作一精确比较,如果完全一致,就认为考生的答案是正确的并依据题目的分值进行给分;否则认为考生答案是错误的而不给相应题目得分。但对主观题的评判过程一般是让考生把其作答的结果形成一个文件(答案文件),再通过网络把答案文件上传到考试服务器中的专用目录中,再通过任课教师在考试结束后对考生的答案文件进行人工评判给分;最后把考生客观题的计算机自动评分结果和主观题的人工评分结果累加起来,作为考生的最终成绩。比如西南科技大学的《数据库应用》课程就是采用这种方法进行的。在考试过程中,客观题的评判由考试系统自动实现,效率较高,不需要人工干扰,而对于主观题,处理的效率就比较低下。

因此文中就此问题提出将人工智能领域中的自然语言理解技术(主要是分词技术)、文本的空间向量模型表示和知识的框架表示内容应用到网络考试系统中的主观题的自动评分过程中。

1 主观题自动评分原理

对于网络考试系统来说,其自动评分是在特定范围内的,不需要让其理解所有的自然语言,只需要理解标准答案即可。因此,应该使用某种算法使标准答案转化成机器能够理解的形式,将考生答案也按照一定的规则转化成计算机可以理解的形式,然后再将其和标准答案进行匹配并评分。其关键问题是,需要将评分规则转化为可以被机器理解的知识库。主观题的自动评分原理见图 1。

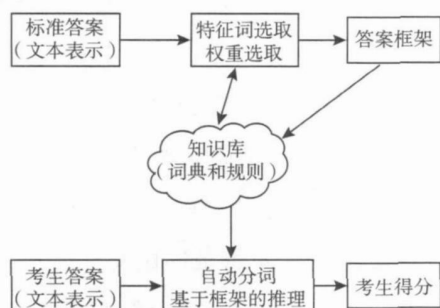


图 1 主观题自动评分原理

Fig 1 Subjective auto marking principles

对于文本信息的智能化处理,制定一个知识库是非常重要的,因为知识库是问题解决方法的关键所在。对自动评分算法而言,知识库的来源主要有 3 个方面:

1) 词典,利用目前在文本信息处理方面经常使用的词典,并且还可以向词典中加入专业词汇以提高数据处理的准确性;

2) 文本特征词及其权重,特征词是一些能够代表文本信息的词的组合,权重是衡量其在文本信息中的重要程度;

3) 文本匹配规则,在本算法中主要检查考生答案和标准答案的匹配程度,因此文本匹配规则也应该输入知识库。

在自动分词环节,本算法在实现时选用最大概率分词算法和最大匹配分词算法相结合;在文本信息处理环节,选用向量空间模型;在知识表示方法中选择框架表示法^[7],亦即通过最大概率分词算法和最大匹配分词算法将考生答案分成汉语的有效词语,再利用这些词语组成考生答案文本的空间向量;根据知识的框架表示法对主观题答案进行知识表示,亦即把标准答案看成是由多个答案要点组成(对应框架表示法中的槽),每个要点又由若干特征词组成(对应框架表示法中的面),并且由有经验的命题教师选定描述标准答案的特征词及指定相应权值,再由系统进行归一化处理后形成评分规则库;最后使用考生答案和评分规则库的标准答案向量表示,计算二者之间的距离(亦即文本的矢量匹配算法),并由此决定考生主观题的最终得分。

2 文本矢量特征匹配算法

主观试题的答案以文本方式存储,而经过分词后的文本如何能更容易地被计算机处理,关系到文本处理的准确性。因此文本表示方法是自动评分算法的关键问题。近年来,在 Web 文本信息特征获取算法的研究中,矢量空间模型(Vector Space Model, VSM)^[8-9]是应用较多且效果较好的方法之一,文中的算法借鉴了该模型的思想。

在矢量空间模型中,文本被看作由一组正交词条所生成的矢量空间。根据这个思想,同时考虑到考试评分中经常将试题答案分为几个要点,因此提出主观题成绩评判模型的步骤如下:

1) 答案文本是由一些要点组成,如果把答案文本(Answer text,用 A 表示)看成一个由 n 个要点(P_i)组成的集合,则可以这样表示答案:

$$A = \{P_1, P_2, \dots, P_i, \dots, P_n\} \quad (1)$$

2) 设每个要点 P_i 的分值为 M_i ,则该答案的总分 M 为

$$M = \sum_{i=1}^n M_i \quad (2)$$

3) 按照 VSM 思想,将标准答案每一个要点 P_i 被看成是由 K_i 个特征词(w_j)组成的向量 P :

$$P = \{w_1, w_2, w_3, \dots, w_j, \dots, w_{K_i}\} \quad (3)$$

4) 设每个特征词的权重是 w_j (由经验丰富的任课教师人工设置), 则其归一化权重为

$$\bar{w}_j = w_j / \sqrt{\sum_{j=1}^{K_i} (w_j)^2} \quad (4)$$

5) 设考生答案的每一个要点 P_i 也被看成是由 K_i 个特征词 (w_j) 组成的向量 P

$$P = \{w_1, w_2, w_3, \dots, w_j, \dots, w_{K_i}\} \quad (5)$$

6) 如果考生答案和标准答案的向量间的距离为 0, 则说明考生答案和标准答案完全匹配, 考生可以拿到该要点的所有分值, 即

$$M_i = (1 - \sqrt{\sum_{j=1}^{K_i} ((w_j - \bar{w}_j) \times \bar{w}_j)^2}) \times M_i \quad (6)$$

7) 根据式 (6), 则考生所得总分 M 为

$$M = \sum_{i=1}^n (1 - \sqrt{\sum_{j=1}^{K_i} ((w_j - \bar{w}_j) \times \bar{w}_j)^2}) \times M_i \quad (7)$$

3 算法测试

文中的系统采用 Oracle 作为后台数据库管理系统, 基于 B/S 模式设计了自动评分测试软件, 其运行界面见图 2。

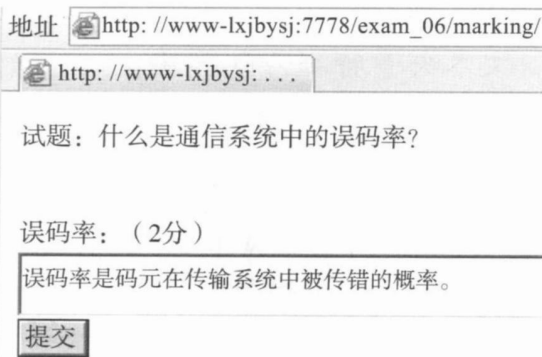


图 2 标准答案自动评分算法测试页面

Fig 2 Test calculating interface of the standard answer

图 2 的界面中输入的答案数据“误码率是码元在传输系统中被传错的概率”是标准答案, 故而其运行结果为得满分 (2 分), 见图 3。如果把考生答案改为“误码率是码元被传错的次数”后, 其答题界面和其计算得分 (1.71 分) 分别见图 4 和图 5。

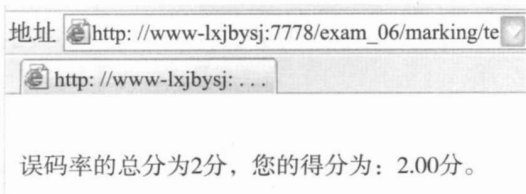


图 3 标准答案自动评分算法计算结果页面

Fig 3 Calculating result page of the standard answer

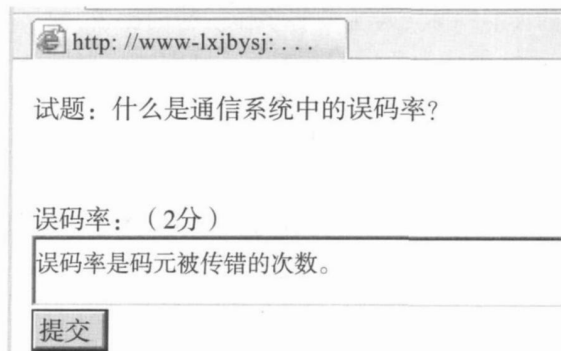


图 4 非标准答案自动评分算法测试页面

Fig. 4 Test calculating interface of the nonstandard answer

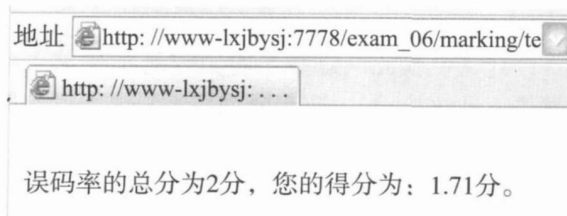


图 5 非标准答案自动评分算法计算结果页面

Fig. 5 Calculating result page of the nonstandard answer

将本算法评得的分数与人工评分相比, 分数的容差在 $(-0.5 \sim +0.5)$, 可以测得其评分的准确度在 86.93%。具体测试情况见表 1。

4 结 语

文中旨在将人工智能关于中文自然语言理解的研究成果 (文本空间向量模型、分词算法) 应用于在线考试的主观题答案的“理解”中, 提出了基于向量空间模型的文本矢量特征匹配算法并应用到主观题的自动评分过程中。实际的数据测试表明: 答案越复杂, 要点越多, 评分的准确性越差; 反之, 要点越少, 答案越简单, 评分的准确性越好。而且人工设置关键词和权重也有利有弊, 人工设置固然增强了系统的准确程度, 但是其前提是设置人必须是有经验的老师, 如果是没有经验的老师设置, 则给算法增加了人为的误差。该算法具有一定的实用性和拓展性, 但还有待进一步完善。

表 1 算法测试结果
Tab 1 Particular test circumstance

题 目	参 考 答 案	准确度 / %	平均准确 度 / %
名词解释: 操作系统	操作系统是在计算机硬件基础上的一个系统软件,是这样一些程序模块的集合(1分) ... 它们管理和控制计算机系统硬件及软件资源(1分),合理地组织计算机工作流程,以便有效地利用这些资源为用户提供一个功能强大、使用方便和可扩展的工作环境(1分),从而在计算机与其他用户之间起到接口的作用(1分)。	80.5	
名词解释: 进程	一个具有独立功能的程序对某个数据集在处理上的动态执行过程(2分),它是资源分配的基本单位(2分)。	87.7	
名词解释: 误码率	误码率是码元在传输系统中被传错的概率。(4分)	92.3	
名词解释: 误信率	误信率是码元的信息量在传输系统中被丢失的概率。(4分)	91.4	86.93
名词解释: 文件	文件是由创建者所定义的一组相关的信息集合(2分)。文件的主要属性包括:文件名,文件类型,文件长度,创建者,创建时间,修改时间,文件定位信息等(2分)。(注:第2个要点写出诸如文件类型、文件目的等也得2分)	86.9	
名词解释: 虚拟内存	虚拟内存是指仅把作业的一部分装入内存便可运行作业的存储器系统(1分);是指具有请求调入和置换功能,能从逻辑上对内存容量进行扩充的一种存储器系统(2分);它的实现毫无例外地都是建立在离散分配存储管理方式的基础上的(1分)。	82.8	

参考文献 (References):

[1] 朱映辉,江玉珍. 计算机自动评卷策略分析与研究 [J]. 电脑知识与技术, 2005 (35): 30-32
ZHU Ying-hui, JIANG Yu-zhen Analysis and research of the policies about auto scoring in computer[J]. Computer Knowledge and Technology, 2005 (35): 30-32 (in Chinese)

[2] 李丁. 计算机考试系统中自动评分策略的研究与实现 [J]. 广东广播电视大学学报, 2002, 11 (4): 30-32
LI Ding Study and realization of auto-scoring tactics in the computer test system [J]. Journal of Guangdong Radio and TV University, 2002, 11 (4): 30-32 (in Chinese)

[3] 张绍阳,武雅丽,谢小军. 计算机文化基础机试自动阅卷研究 [J]. 现代电子技术, 2003 (1): 90-92
ZHANG Shao-yang, WU Ya-li, XIE Xiao-jun Studying of grading the computer culture foundation exam on computer [J]. Modern Electronic Technique, 2003 (1): 90-92 (in Chinese)

[4] 黄安健. 实现无纸化考试的二项技术处理 [J]. 上海应用技术学院学报, 2003, 3 (1): 66-68
HUANG An-jian Two kinds of technological processing of the realization of the nonpaperlized examination [J]. Journal of Shanghai Institute of Technology, 2003, 3 (1): 66-68 (in Chinese)

[5] 毕蕊. 自动组卷及评分系统的设计 [J]. 中国电化教育, 2002 (2): 84-85.
HUA Rui Auto set the design of book and grade point system [J]. China Educational Technology, 2002 (2): 84-85. (in Chinese)

[6] 李少芳,车艳. 近似串匹配算法在自动评分系统中的应用 [J]. 东莞理工学院学报, 2008, 15 (3): 25-28
LI Shao-fang, SHE Yan The application of approximate string matching algorithms in the auto scoring system [J]. Journal of Dongguan University of Technology, 2008, 15 (3): 25-28 (in Chinese)

[7] 于中华,唐常杰,张天庆. 自然语言句法结构的框架树表示方法 [J]. 小型微型计算机系统, 1999, 20 (8): 583-586
YU Zhong-hua, TANG Chang-jie, ZHANG Tian-qing Tree representation for syntax structure of natural languages[J]. Mini-Micro Systems, 1999, 20 (8): 583-586 (in Chinese)

[8] Sowa J F. Conceptual Structure: Information Processing in Mind and Machine [M]. Boston : Addison Wesley Publishing Co Inc, 1984.

[9] Raghavan V, Wong SKM. A critical analysis of vector space model for information retrieval[J]. Journal of the American Society for Information Science, 1986, 37 (5): 279-287.

(责任编辑:杨 勇,彭守敏)